**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/475, 16/22, A61K 38/18 — A1

(11) International Publication Number: **WO 99/01553**

(43) International Publication Date: 14 January 1999 (14.01.99)

(21) International Application Number: PCT/US98/11462

(22) International Filing Date: 3 June 1998 (03.06.98)

(30) Priority Data:
08/887,997 3 July 1997 (03.07.97) US

(71) Applicants: GENETICS INSTITUTE, INC. [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US). THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 22nd floor, 300 Lakeside Drive, Oakland, CA 94612 (US).

(72) Inventors: FOLLETTIE, Maximillian; 187 Common Street, Belmont, MA 02178 (US). DeROBERTIS, Edward, M.; 16958 Dulce Ynez Lane, Pacific Palisades, CA 90272 (US).

(74) Agent: LAZAR, Steven, R.; Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*

(54) Title: MURINE AND HUMAN CERBERUS–LIKE PROTEINS AND COMPOSITIONS COMPRISING THEM

(57) Abstract

Purified mammalian *cerberus* proteins and process for producing them are disclosed. DNA molecules encoding the mammalian *cerberus* proteins are also disclosed. The proteins may be used for inducing formation, growth, differentiation, proliferation and/or maintenance of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, and for other tissue repair, including cardiac and endoderm.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

## TITLE OF THE INVENTION

MURINE AND HUMAN CERBERUS-LIKE PROTEINS AND COMPOSITIONS COMPRISING THEM

The present invention relates to novel members of the *cerberus* protein family, DNA encoding them, and processes for obtaining them. These proteins may be used to induce expression of factors in and/or differentiation of tissue and organs, and particularly, inducing formation, growth, differentiation, proliferation and/or maintenance of neural, endoderm and cardiac tissue. Thus, these proteins may be useful in the treatment of wounds, tumors, and in the enhancement and/or inhibition of cellular formation, growth, differentiation, proliferation and/or maintenance of other tissue and organs, for example, epidermal, pancreatic, liver, spleen, lung, kidney, brain and/or other tissue. These proteins may also be used for augmenting the activity of other tissue regenerating and differentiation factors, such as the BMPs. The protein has been named mammalian *cerberus-like* by the inventors.

## BACKGROUND OF THE INVENTION

The search for the molecule or molecules responsible for the formation, proliferation, differentiation and maintenance of tissue and organs, such as neurons and related neuronal cells and tissues, has been extensive as there is a tremendous need for factors useful for treating conditions involving degradation or damage to these tissues. A *Xenopus* protein previously identified in embryos, *cerberus*, appears to be involved in induction of the head. Bouwmeester et al., Nature, 382:595-601 (1996).

## SUMMARY OF THE INVENTION

The inventors herein have discovered novel mammalian members of the *cerberus* family of proteins and have surprisingly discovered that members of the *cerberus* protein family are able to induce, enhance and/or inhibit the formation, growth, proliferation, differentiation, maintenance of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes. Accordingly, the present invention provides methods for inducing formation of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, comprising administering to progenitor cells

a composition comprising at least one protein which is a member of the *cerberus* protein family.

The present invention relates to a family of proteins designated as *cerberus*, which appears to be a pioneer protein, with a 9 cysteine residue pattern, which is present in the embryo. In *Xenopus*, *cerberus* mRNA is expressed at low levels in the unfertilized egg, and zygotic transcripts start accumulating at early gastrula. Expression continues during gastrulation and early neurulation, rapidly declining during neurulation. Importantly, *cerberus* expression starts about one hour after that of chordin, suggesting that *cerberus* could act downstream of the chordin signal. The *cerberus* domain of the organizer includes the leading edge of the most anterior organizer cells and extends into the lateral mesoderm. The leading edge gives rise to liver, pancreas and foregut in its midline, and the more lateral region gives rise to heart mesoderm at later stages of development.

In preferred embodiments, the composition may comprise a protein having the amino acid sequence of SEQUENCE ID NO:2 beginning at amino acid 1, 18 to 24, 41, 85 to 91 or 162, and ending at amino acid 241 or 272; or SEQUENCE ID NO:8 beginning at amino acid 1, 18 to 25, 41, 85 to 91 or 162, and ending at amino acid 241 or 267. In one embodiment, the method comprises administering the composition to a patient *in vivo*. Alternatively, the method may comprise administering the composition to cells *in vitro* and recovering neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, which may subsequently be administered to a patient. The composition may further comprise a suitable carrier for administration.

The present invention also provides novel DNA sequences encoding novel members of the *cerberus* protein family. In particular embodiments, the present invention provides novel DNA sequences encoding mammalian *cerberus* proteins such as murine and human *cerberus*. The nucleotide sequences, and the corresponding amino acid sequences encoded by these DNA sequences, are provided in the Sequence Listings. In particular, the present invention comprises isolated DNA sequence encoding a mammalian *cerberus* protein comprising a DNA sequence selected from the group consisting of: nucleotides beginning at #58, 109, 178, 313, 316, 319, 322, 325, 328, or 541 and ending at #780 or 873 of SEQ ID NO: 1; nucleotides beginning at # 1, 52, 55, 58, 61, 64, 67, 70, 73, 121, 256, 259, 262, 265, 268, 271 or 484 and ending at #723 or

801 of SEQ ID NO:7; or nucleotides encoding amino acids beginning at #1, 18, 41, 85 to 91 or 162 and ending at #241 or 272 of SEQ ID NO: 2; or amino acids beginning at # 1, 18 to 25, 41, 85 to 91 or 162 and ending at #241 or 267 of SEQ ID NO: 8, as well as fragments and variants of the above sequences which are readily obtainable from the above and which maintain *cerberus* activity. The present invention further comprises sequences which hybridize to these sequences under stringent hybridization conditions and encode a protein which exhibits *cerberus* activity.

It is expected that mammalian *cerberus* protein, as expressed by mammalian cells such as CHO cells, exists as a heterogeneous population of active species of *cerberus* protein with varying N-termini. Based in part upon the Von Heginje signal peptide prediction algorithm, the first 17 to 24 amino acids appear to be involved in signaling for the secretion of the mature peptide. It is expected that active species may optionally include the signal peptide and will include amino acid sequences beginning with amino acid residues #1, 18, 19, 20, 21, 22, 23, 24 or 25 of SEQ ID NO:2 or SEQ ID NO:8. Thus, it is expected that DNA sequences encoding active mammalian *cerberus* proteins include those comprising nucleotides #109, 112, 115, 118, 121, 124, 127 or 130 to #780 or 873 of SEQ ID NO: 1; or comprising nucleotides #1, 52, 55, 58, 61, 64, 67, 70 or 73 to #723 or 801 of SEQ ID NO:7. Accordingly, active species of *cerberus-like* protein are expected to include those comprising amino acids #1, 18, 19, 20, 21, 22, 23, 24 or 25 to #241 or 272 of SEQ ID NO:2; or amino acids # 1, 18, 19, 20, 21, 22, 23, 24 or 25 to #241 or 267 of SEQ ID NO:8.

As described further herein, it is further expected that *cerberus* and *cerberus-like* proteins may be proteolytically processed by cells to form further active species. For example, putative proteolytic processing sites for cleavage, which are typically characterized by the formula R-X-K/R-R, are found at amino acids 37 to 40 and 82 to 85 of SEQ ID NO:2 or SEQ ID NO:8. Thus, it is expected that DNA sequences encoding active mammalian *cerberus* proteins include those comprising nucleotides #178 or 313 to #780 or 873 of SEQ ID NO:1; and # 121 or 256 to #723 or 801 of SEQ ID NO:7. Accordingly, further active species of *cerberus-like* protein are expected to include those comprising amino acids beginning at #41 or 86 and ending at #241 or 272 of SEQ ID

NO: 2; or comprising amino acids beginning at #41 or 86 and ending at #241 or 267 of SEQ ID NO:8.

The above sequences of SEQ ID NO:1 and 2 are used to isolate and sequence the human *Cerberus*-like DNA and amino acid sequences.

In yet another embodiment, the present invention comprises a method of altering the regulation of genes in a patient in need of same comprising administering to said patient an effective amount of the above compositions. The alteration of regulation of neuronal genes may be accomplished by stimulating or inhibiting binding by *cerberus* proteins of receptor proteins, including bone morphogenetic proteins [BMPs]. Thus, the mammalian *cerberus* and *cerberus-like* protein family may be capable of inducing formation of neural, bone, cartilage and other tissue. The mammalian *cerberus* and *cerberus-like* protein family may also be capable of inhibiting, augmenting or otherwise affecting the activity of its receptor molecules to which it binds, including molecules of the BMP family.

In other embodiments, the present invention comprises vectors comprising the above DNA molecules in operative association with an expression control sequence therefor, as well as host cells transformed with these vectors. In yet other embodiments, the present invention comprises methods for producing purified mammalian *cerberus* proteins, novel mammalian *cerberus* proteins, and compositions containing the mammalian *cerberus* proteins. These methods may comprise the steps of: culturing a host cell transformed with a DNA sequence encoding a mammalian *cerberus* protein such as described above; and recovering and purifying said mammalian *cerberus* protein from the culture medium. The present invention further comprises the purified mammalian *cerberus* polypeptide produced by the above methods, as well as purified mammalian *cerberus* polypeptides comprising an amino acid sequence encoded by the above DNA sequences. The proteins of the present invention may comprise the amino acid sequence beginning at amino acid #1, 18, 19, 20, 21, 22, 23, 24, 25, 86, 87, 88, 89, 90, 91 or 162 and ending at #241 or 272 of SEQ ID NO:2; #1, 18, 19, 20, 21, 22, 23, 24, 25, 86, 87, 88, 89, 90, 91 or 162 and ending at #241 or 267 of SEQ ID NO:8; or a mammalian *cerberus* protein having a molecular weight of about 20-30 kd, said protein comprising an amino acid sequence highly homologous to the amino acid sequences of SEQ ID NO:2

or SEQ ID NO:8, and having the ability to regulate the transcription of one or more genes. One species of active *cerberus-like* protein is a mature peptide contemplated to comprise the amino acid sequence from amino acids 18 to 272 of SEQ ID NO:2; or amino acid 18 to 267 of SEQ ID NO:8, each expected to have a molecular weight of about 28.6 kD. Another species of active *cerberus-like* protein is a cleaved peptide contemplated to comprise the amino acid sequence from amino acids 86 to 91 and ending at amino acid 272 of SEQ ID NO:2, particularly from 18 or 90 to 272 of SEQ ID NO:2, or from amino acids 86 to 91 ending at amino acid 267 of SEQ ID NO:8, particularly from 18 or 90 to 267 of SEQ ID NO:8. The mature polypeptide for each of these proteins is expected to have a molecular weight of about 20.7 kD.

**Description of Sequences**

SEQ ID NO: 1 nucleotide sequence of mammalian *cerberus* DNA, particularly murine *cerberus* DNA.

SEQ ID NO: 2 amino acid sequence of the mammalian *cerberus* protein encoded by SEQ ID NO: 1.

SEQ ID NO: 3 to 5 are consensus nucleotide sequences of probes to the *cerberus* and *cerberus-like* proteins.

SEQ ID NO: 6 is the genomic DNA sequence encoding human *Cerberus-like* protein. The symbol "N" indicates that the nucleotide residue may be any of A, C, T or G.

SEQ ID NO: 7 is the cDNA sequence encoding human *Cerberus-like* protein.

SEQ ID NO: 8 is the amino acids sequence of human *Cerberus-like* protein encoded by SEQ ID NO: 7.

**Description of ATCC Deposits**

An *E. coli* DH5α strain transformed with pGIMCerb, which comprises the mammalian *cerberus* coding sequence described in Sequence ID NO: 1 has been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852, and has been accorded the ATCC accession number 98347.

**Detailed Description of the Invention**

As used herein, the terms "*cerberus*" or "*cerberus-like*" are both used to signify the protein family which comprises the *cerberus* and *cerberus-like* proteins. "*cerberus* or *cerberus-like* protein" refers to mammalian *cerberus* and *cerberus-like* proteins, such

as the murine or human *cerberus* proteins, and other proteins which share sequence homology to the highly conserved cysteine pattern of the C-terminal portion of the mammalian *cerberus* proteins. One specific member of the *cerberus* protein family is the murine *cerberus-like* protein, having the amino acid sequence specified in SEQUENCE ID NO:2, as well as homologues of this protein found in other species; and other proteins which are closely related structurally and/or functionally to murine *cerberus*. It is also expected that *cerberus* related proteins also exist in other species, including family members in *Xenopus*, and *Drosophila*. *C. elegans*, zebrafish, as well as in all mammals, for example, rats, mice and humans. "*Cerberus* or *cerberus-like* proteins" also includes variants of the *cerberus* proteins, such as allelic variants or variants induced by mutagenesis or deletions, and fragments of *cerberus* proteins which variants and fragments retain *cerberus* activity. "*Cerberus* and *Cerberus-like*" are also used to signify the family of proteins sharing structural and/or functional similarity, including those proteins which are described further herein.

As used herein, the term "*cerberus* or *cerberus-like* activity" refers to one or more of the activities which are exhibited by the mammalian *cerberus-like* proteins of the present invention. In particular, "*cerberus* or *cerberus-like* activity" includes the ability to induce, enhance and/or inhibit the formation, growth, proliferation, differentiation, maintenance of neurons and/or related neural cells and tissues such as brain cells, Schwann cells, glial cells and astrocytes. "*Cerberus* or *cerberus-like*" activity also includes the ability to induce molecular markers of neuroendocrine or ectoderm tissue, such as *OTX2*, N-CAM, MASH, chromagranin, and AP2, as well as the ability to induce the formation of neurons and/or related neural cells and tissues such as brain cells, Schwann cells, glial cells and astrocytes. "*Cerberus* or *cerberus-like* activity" may also include the ability to regulate the interaction of ligands and their protein receptors. For example, "*cerberus* or *cerberus-like* activity" may include the ability to bind to one or more members of the bone morphogenetic protein [BMP] and/or *wnt* protein families, and thereby inhibit, augment or otherwise affect the activity of such molecules.

"*Cerberus* or *cerberus-like* activity" may further include the ability to regulate the formation, differentiation, proliferation and/or maintenance of other cells and/or tissue, for example connective tissue, organs and wound healing. In particular, "*cerberus* or *cerberus-like* activity" may include the ability to enhance and/or inhibit the formation, growth, proliferation, differentiation and/or maintenance of cardiac, spleen, liver, pancreas, stomach, kidney, lung and brain cells and tissue, as well as osteoblasts and bone, chondrocytes and cartilage, tendon, epidermis and muscle. "*Cerberus* and *cerberus-like* activity" also includes the activities of *cerberus* and *cerberus-like* protein in the assays described in the examples and specification herein.

*Cerberus* and *cerberus-like* cDNA should be useful as a diagnostic tool (such as through use of antibodies in assays for proteins in cell lines or use of oligonucleotides as primers in a PCR test to amplify those with sequence similarities to the oligonucleotide primer, and to determine how much *cerberus* is present). *Cerberus* might act upon its target cells via its own receptor. *Cerberus*, therefore, may be useful for the isolation of that receptor. In addition, *cerberus* or its receptor should prove useful as a diagnostic probe for certain tumor types. Thus, *cerberus*, its receptor, or antibodies to either may be potent agonists or antagonists which may be clinically useful. In addition, complexes of *cerberus* and its receptor, *cerberus* and antibodies to it, or *cerberus* receptor and antibodies to it, may each be useful in a number of *in vitro*, *ex vivo* or clinical uses.

The present invention also includes protein variants and functional fragments of the amino acid sequence of the mammalian *cerberus* protein shown in SEQ ID NO: 2 which retain *cerberus* activity. The present invention also includes antibodies to a purified mammalian *cerberus* protein such as the above. The compositions of the present invention comprise a therapeutic amount of at least one of the above mammalian *cerberus* proteins. It is expected that such protein variants and functional fragments of *cerberus* or *cerberus-like* proteins will include amino acid sequences which share significant homology with the amino acid sequence of SEQ ID NO: 2, most preferably at least 80% or 90% amino acid identity. The variants and functional fragments which retain *cerberus-like* activity are expected to include those which retain the cysteine pattern found in the SEQ ID NO: 2. For example, a truncated polypeptide comprising amino acids #162 to #241 of SEQ ID NO: 2; or from #162 to #241 of SEQ ID NO:8 will

each retain the full 9 cysteine pattern found in the carboxy terminal portion of the *cerberus* and *cerberus-like* sequences of SEQ ID NO: 2 and SEQ ID NO:8, respectively.

In yet another embodiment, the present invention comprises a method of altering the regulation of genes in a patient in need of same comprising administering to said patient an effective amount of the above compositions. For example, the alteration of regulation of neuronal genes may be accomplished by stimulating or inhibiting binding of receptor proteins, for example, binding between the mammalian *cerberus* protein and its receptor protein, such as a *wnt* protein, or a BMP protein. Thus, *cerberus* proteins may be capable of regulating the binding interaction of ligands to their receptor proteins, as well as the interaction of transcriptional factors on cells.

The present invention also encompasses hybrid or fusion vectors comprising the coding DNA sequences of the present invention and other *cerberus* encoding sequences, linked to a tissue specific or inducible regulatory sequence, such as a promoter or operator. In a preferred embodiment of the invention, the coding sequence for mammalian *cerberus-like* protein is operably linked to one or more promoters, enhancers and/or other regulatory elements from genes which are selectively expressed in neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes. For example, the promoter of the GFAP gene, which is known to be expressed in astrocytes and neuronal cells; and the promoter of the *OTX2* gene, which is known to be expressed in the anterior brain, are suitable for the tissue specific production of *cerberus*. Additionally, the DNA sequence encoding mammalian *cerberus* may be operatively linked to one or more regulatory sequences from GFAP or *OTX2* proteins, as well as other proteins which are selectively produced in neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes.

In other preferred embodiments of the invention, the coding sequence for mammalian *cerberus-like* protein is operably linked to the promoter isolated from other genes, organs or cells of interest. Vectors using other tissue-selective regulatory elements and inducible regulatory elements may also be useful for the selective or inducible expression of the mammalian *cerberus-like* proteins of the present invention.

Another aspect of the invention provides pharmaceutical compositions containing a therapeutically effective amount of mammalian *cerberus-like* protein, in a

pharmaceutically acceptable vehicle or carrier. These compositions of the invention may be used in the formation of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, as well as liver, pancreas, lung, heart, kidney, spleen, stomach, cardiac tissue and cells, as well as connective tissue and cells, including osteocytes, chondrocytes, myocytes, tendon cells, epidermal cells, and adipocytes. These compositions may further be utilized in order to enhance and/or inhibit the formation, growth, proliferation, differentiation and/or maintenance of bone, osteoblasts, cartilage, chondrocytes, beta cells and other cell types typically found in the islets of Langerhans or other pancreatic cells, as well as other organ tissues such as epidermis, spleen, brain, lung and kidney tissue. The compositions comprising mammalian *cerberus-like* protein may be used to treat precursor or stem cells, such as endoderm cells, which are able to differentiate into cells which comprise differentiated tissue or organs, such as cardiac and neural cells, in order to enhance the formation, differentiation, proliferation and/or maintenance of such cells, tissue or organs. Methods for forming and maintaining pancreatic cells are described, for example, in WO93/00441, the disclosure of which is hereby incorporated herein by reference. In addition, the compositions may be used to regulate embryonic development, for instance, by affecting the development of embryonic cells and tissue into the endodermal phenotype.

The compositions of the invention may comprise, in addition to a mammalian *cerberus-like* protein, other therapeutically useful agents including growth factors such as epidermal growth factor (EGF), fibroblast growth factor (FGF), transforming growth factor (TGF-α and TGF-β), *Wnts*, *hedgehogs*, including sonic, indian and desert *hedgehogs*, activins, inhibins, bone morphogenetic proteins (BMP), and insulin-like growth factor (IGF). The compositions may also include an appropriate matrix, for instance, for supporting the composition and providing a surface for ingrowth of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, or for other tissue or cell growth. The matrix may provide slow release of the mammalian *cerberus-like* protein and/or the appropriate environment for presentation thereof.

The mammalian *cerberus-like* protein containing compositions may be employed in methods for treating a number of tissue defects, and healing and maintenance of

various types of tissues and wounds. The tissues and wounds which may be treated include repair or induction of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes. It also includes treatment of cardiac, liver, pancreas, spleen, lung, kidney, brain and stomach tissue and may also include cartilage, epidermis, muscle, including cardiac muscle, other connective tissue, such as bone, tendon and ligament and other tissues and wounds. These methods, according to the invention, entail administering to a patient needing such tissue formation, wound healing or tissue repair, an effective amount of mammalian *cerberus* protein. The mammalian *cerberus-like* containing compositions may also be used to treat or prevent degenerative nerve conditions such as Parkinson's Disease, Alzheimer's Disease and Lou Gehrig's Disease, as well as other degenerative nerve diseases, and other conditions involving defects of neural tissue. The compositions may also be useful to treat other conditions such as osteoporosis, rheumatoid arthritis, osteoarthritis, and other abnormalities of connective tissue, or of other organs or tissues, such as muscle, pancreas, liver, spleen, lung, cardiac, brain, and kidney tissue, and other tissues and organs. These methods may also entail the administration of a protein of the invention in conjunction with administration of at least one other protein, for example growth factors including EGF, FGF, TGF-α, TGF-β, BMP, *Wnts*, *hedgehogs*, including sonic, indian and desert *hedgehogs*, activin, inhibin and IGF. In a particular embodiment of the present invention the mammalian *cerberus-like* gene or protein may be used to augment the activities of BMPs or other members of the TGF-β superfamily.

Still a further aspect of the invention are DNA sequences coding for expression of mammalian *cerberus-like* protein. Such sequences include the sequence of nucleotides in a 5' to 3' direction illustrated in SEQ ID NO:1 or SEQ ID NO:7, DNA sequences which, but for the degeneracy of the genetic code, are identical to the DNA sequence SEQ ID NO:1 or SEQ ID NO:7, and encode the protein of SEQ ID NO:2 or SEQ ID NO:8. Further included in the present invention are DNA sequences which hybridize under stringent conditions with the DNA sequence of SEQ ID NO:1 or SEQ ID NO:7 and encode a protein having *cerberus-like* activity. Preferred DNA sequences include those which hybridize under stringent conditions [see, T. Maniatis et al, Molecular Cloning (A Laboratory Manual), Cold Spring Harbor Laboratory (1982), pages 387 to 389]. It is

generally preferred that such DNA sequences encode a polypeptide which is at least about 80% homologous, and more preferably at least about 90% homologous, to the mature mammalian *cerberus-like* amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:8. Further, allelic or other variations of the sequences of SEQ ID NO: 1 or SEQ ID NO:2, whether such nucleotide changes result in changes in the peptide sequence or not, but where the peptide sequence still has *cerberus-like* activity, are also included in the present invention. The present invention also includes functional fragments of the DNA sequence of mammalian *cerberus-like* proteins shown in SEQ ID NO: 1 or SEQ ID NO:7 which encode a polypeptide which retains the activity of *cerberus-like* protein. The determination whether a particular variant or fragment of the mammalian *cerberus-like* protein of the present invention, such as those shown in SEQ ID NO:2 or SEQ ID NO:8, maintain *cerberus-like* activity, is routinely performed using the assays described in the examples and specification herein.

The DNA sequences of the present invention are useful, for example, as probes for the detection of mRNA encoding other *cerberus-like* proteins in a given cell population. The DNA sequences may also be useful for preparing vectors for gene therapy applications as described below.

A further aspect of the invention includes vectors comprising a DNA sequence as described above in operative association with an expression control sequence therefor. These vectors may be employed in a novel process for producing a recombinant mammalian *cerberus-like* protein of the invention in which a cell line transformed with a DNA sequence encoding mammalian *cerberus-like* protein in operative association with an expression control sequence therefor, is cultured in a suitable culture medium and mammalian *cerberus-like* protein is recovered and purified therefrom. This process may employ a number of known cells both prokaryotic and eukaryotic as host cells for expression of the polypeptide. The vectors may also be used in gene therapy applications. In such use, the vectors may be transfected into the cells of a patient *ex vivo*, and the cells may be reintroduced into a patient, or the vectors may be introduced into a patient *in vivo* through targeted transfection. Alternatively, homologous *cerberus* gene expression may be upregulated by known recombination techniques to insert high expression regulatory elements into the genome in proximity to the *cerberus* coding sequence disclosed herein.

In a preferred embodiment of the invention, vectors are prepared using one or more non-native regulatory elements, such as promoters and/or enhancers operatively associated with the coding sequence for mammalian *cerberus-like*, in order to achieve expression of mammalian *cerberus-like* in desired cell tissue and/or at a desired time in development. For example, a vector may be constructed using the promoter element from genes, which is known to be constitutively expressed in neuronal development. By operatively associating the promoter from suitable genes with the coding sequence for *cerberus-like*, and transforming suitable cells, such as neuronal stem cells, one can express mammalian *cerberus-like* in these cells, thus promoting the desired effects of formation, growth, proliferation, differentiation and/or maintenance of cells such as neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, either in *in vitro* culture or *in vivo*.

Still a further aspect of the invention are mammalian *cerberus-like* proteins or polypeptides. Such polypeptides are characterized by having an amino acid sequence including the sequence illustrated in SEQ ID NO:2 or SEQ ID NO:8, variants of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:8, including naturally occurring allelic variants, and other variants in which the protein retains *cerberus-like* activity, for example, the ability to enhance and/or inhibit the formation, growth, proliferation, differentiation and/or maintenance of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, and may also affect the formation, growth, proliferation, differentiation and/or maintenance of pancreas, liver, stomach, cardiac, or other tissue such as bone, osteocytes, chondrocytes and/or cartilage tissue, or other organ tissue, such as spleen, lung, brain and kidney tissue, characteristic of *cerberus-like* protein. Preferred polypeptides include a polypeptide which is at least about 80% and more preferably at least about 90% homologous to the mature mammalian *cerberus-like* amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:8. Further, allelic or other variations of the sequences of SEQ ID NO:2 or SEQ ID NO:8, whether such amino acid changes are induced by mutagenesis, chemical alteration, or by alteration of DNA sequence used to produce the polypeptide, where the peptide sequence still has *cerberus-like* activity, are also included in the present invention. The present invention also includes fragments of the amino acid sequence of mammalian *cerberus-like* shown in

SEQ ID NO:2 or SEQ ID NO:8 which retain the activity of *cerberus-like* protein. One skilled in the art can readily produce such variations and fragments of mammalian *cerberus-like* protein using techniques known in the art, and can readily assay them for activity, as described herein.

The purified proteins of the present inventions may be used to generate antibodies, either monoclonal or polyclonal, to mammalian *cerberus-like* proteins and/or other related proteins, using methods that are known in the art of antibody production. Thus, the present invention also includes antibodies to mammalian *cerberus* and/or other *cerberus-like* proteins. The antibodies may be useful for purification of mammalian *cerberus-like* proteins, or for inhibiting or preventing the effects of *cerberus* proteins either *in vitro* or *in vivo*. The mammalian *cerberus-like* proteins may be useful for inducing the growth and/or differentiation of embryonic cells and/or stem cells. Thus, the proteins or compositions of the present invention may also be useful for treating cell populations, such as embryonic cells or stem cell populations, to enhance, enrich or to inhibit the growth and/or differentiation of the cells. For example, the mammalian *cerberus-like* proteins may be useful for treating cell populations to enhance and/or inhibit the formation, differentiation, proliferation and/or maintenance of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes and/or other cells and tissue. The treated cell populations may be useful for, among other things, gene therapy applications, as described below. Thus, the proteins of the present invention may be useful in wound healing, tissue and organ repair and regeneration processes, as well as in differentiation of tissue, for example in embryonic development. In particular, it has been observed by the inventors that the mammalian *cerberus-like* protein may be useful for the induction, formation, growth, differentiation, proliferation and/or maintenance and repair of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes. The *cerberus-like* proteins are normally present as secreted proteins, and have been demonstrated to have effects on the growth and differentiation of neuronal and other neural cell and tissue types. Thus, these proteins, and compositions containing them, may be useful in the treatment of nerve and brain disorders, such as Parkinson's disease, Alzheimer's disease, and in the enhancement and/or inhibition of cellular formation, growth, differentiation, proliferation and/or

maintenance, for example formation of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes.

The mammalian *cerberus-like* proteins provided herein include factors encoded by the sequences similar to those of SEQ ID NO:1 or SEQ ID NO:7, but into which modifications or deletions are naturally provided (e.g. allelic variations in the nucleotide sequence which may result in amino acid changes in the polypeptide) or deliberately engineered. For example, synthetic polypeptides may wholly or partially duplicate continuous sequences of the amino acid residues of SEQ ID NO:2 or SEQ ID NO:8. These sequences, by virtue of sharing primary, secondary, or tertiary structural and conformational characteristics with mammalian *cerberus-like* polypeptides of SEQ ID NO:2 or SEQ ID NO:8 may possess biological properties in common therewith. Thus, these modifications and deletions of the native mammalian *cerberus-like* may be employed as biologically active substitutes for naturally-occurring mammalian *cerberus-like* polypeptides in therapeutic processes. It can be readily determined whether a given variant or fragment of mammalian *cerberus-like* protein maintains the biological activity of *cerberus* by subjecting both mammalian *cerberus-like* and the variant or fragment of mammalian *cerberus-like* to the assays described herein.

Other specific mutations of the sequences of mammalian *cerberus-like* proteins described herein involve modifications of glycosylation sites. These modifications may involve O-linked or N-linked glycosylation sites. For instance, the absence of glycosylation or only partial glycosylation results from amino acid substitution or deletion at asparagine-linked glycosylation recognition sites. The asparagine-linked glycosylation recognition sites comprise tripeptide sequences which are specifically recognized by appropriate cellular glycosylation enzymes. These tripeptide sequences are either asparagine-X-threonine or asparagine-X-serine, where X is usually any amino acid. A variety of amino acid substitutions or deletions at one or both of the first or third amino acid positions of a glycosylation recognition site (and/or amino acid deletion at the second position) results in non-glycosylation at the modified tripeptide sequence. Such variants of mammalian *cerberus-like* are within the present invention. Additionally, bacterial expression of mammalian *cerberus-like* proteins will result in production of a non-glycosylated protein, even if the glycosylation sites are left unmodified. Such

bacterially produced versions of mammalian *cerberus-like* are within the present invention.

The present invention also encompasses the novel DNA sequences, free of association with DNA sequences encoding other proteinaceous materials, and coding for expression of mammalian *cerberus-like* proteins. These DNA sequences include those depicted in SEQ ID NO:1 or SEQ ID NO:7 in a 5' to 3' direction and those sequences which hybridize thereto under stringent hybridization conditions [for example, 0.1X SSC, 0.1% SDS at 65°C; see, T. Maniatis et al, Molecular Cloning (A Laboratory Manual), Cold Spring Harbor Laboratory (1982), pages 387 to 389] and encode a protein having *cerberus-like* activity. Stringent hybridization conditions also refer to initial low stringency hybridization conditions, followed by higher stringency wash conditions. These DNA sequences also include those which comprise variants and fragments of the DNA sequence of SEQ ID NO:1 SEQ ID NO:7 which hybridize thereto under stringent hybridization conditions and encode a protein having *cerberus-like* activity.

Similarly, DNA sequences which code for mammalian *cerberus-like* proteins coded for by the sequences of SEQ ID NO:1 or SEQ ID NO:7, or mammalian *cerberus-like* proteins which comprise the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:8, but which differ in codon sequence due to the degeneracies of the genetic code or allelic variations (naturally-occurring base changes in the species population which may or may not result in an amino acid change) also encode the novel factors described herein. Variations in the DNA sequences of SEQ ID NO:1 or SEQ ID NO:7 which are caused by point mutations or by induced modifications (including insertion, deletion, and substitution) to enhance the activity, half-life or production of the polypeptides encoded are also encompassed in the invention.

Another aspect of the present invention provides a novel method for producing mammalian *cerberus-like* proteins. The method of the present invention involves culturing a suitable cell line, which has been transformed with a DNA sequence encoding a mammalian *cerberus-like* protein of the invention, under the control of known regulatory sequences. The transformed host cells are cultured and the mammalian *cerberus-like* proteins recovered and purified from the culture medium. The purified

proteins are substantially free from other proteins with which they are co-produced as well as from other contaminants.

Suitable cells or cell lines may be mammalian cells, such as Chinese hamster ovary cells (CHO). The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening, product production and purification are known in the art. See, e.g., Gething and Sambrook, Nature, 293:620-625 (1981), or alternatively, Kaufman et al, Mol. Cell. Biol., 5(7):1750-1759 (1985) or Howley et al, U.S. Patent 4,419,446. Another suitable mammalian cell line, which is described in the accompanying examples, is the monkey COS-1 cell line. The mammalian cell CV-1 may also be suitable.

Bacterial cells may also be suitable hosts. For example, the various strains of E. coli (e.g., HB101, MC1061) are well-known as host cells in the field of biotechnology. Various strains of B. subtilis, Pseudomonas, other bacilli and the like may also be employed in this method. For expression of the protein in bacterial cells, DNA encoding the propeptide of *cerberus-like* is generally not necessary.

Many strains of yeast cells known to those skilled in the art may also be available as host cells for expression of the polypeptides of the present invention. Additionally, where desired, insect cells may be utilized as host cells in the method of the present invention. See, e.g. Miller et al, Genetic Engineering, 8:277-298 (Plenum Press 1986) and references cited therein.

Another aspect of the present invention provides vectors for use in the method of expression of these novel mammalian *cerberus-like* polypeptides. Preferably the vectors contain the full novel DNA sequences described above which encode the novel factors of the invention. Additionally, the vectors contain appropriate expression control sequences permitting expression of the *cerberus-like* protein sequences. Alternatively, vectors incorporating modified sequences as described above are also embodiments of the present invention. Additionally, the sequence of SEQ ID NO:1, SEQ ID NO:7 or other sequences encoding mammalian *cerberus-like* proteins could be manipulated to express a mature mammalian *cerberus-like* protein by deleting mammalian *cerberus-like* propeptide sequences and replacing them with sequences encoding the complete propeptides of other *cerberus-like* proteins or other suitable propeptides. Thus, the

present invention includes chimeric DNA molecules encoding a propeptide from a member of the *cerberus-like* family linked in correct reading frame to a DNA sequence encoding a mammalian *cerberus-like* polypeptide.

The vectors may be employed in the method of transforming cell lines and contain selected regulatory sequences in operative association with the DNA coding sequences of the invention which are capable of directing the replication and expression thereof in selected host cells. Regulatory sequences for such vectors are known to those skilled in the art and may be selected depending upon the host cells. Such selection is routine and does not form part of the present invention.

In order to produce rat, human or other mammalian *cerberus-like* proteins, the DNA encoding it is transferred into an appropriate expression vector and introduced into mammalian cells or other preferred eukaryotic or prokaryotic hosts by conventional genetic engineering techniques. The preferred expression system for biologically active recombinant mammalian *cerberus-like* is contemplated to be stably transformed mammalian cells.

One skilled in the art can construct mammalian expression vectors by employing the sequence of SEQ ID NO:1, SEQ ID NO:7, or other DNA sequences encoding *cerberus-like* proteins or other modified sequences and known vectors, such as pCD [Okayama et al., Mol. Cell Biol., 2:161-170 (1982)], pJL3, pJL4 [Gough et al., EMBO J., 4:645-653 (1985)] and pMT2 CXM.

The mammalian expression vector pMT2 CXM is a derivative of p91023(b) (Wong et al., Science 228:810-815, 1985) differing from the latter in that it contains the ampicillin resistance gene in place of the tetracycline resistance gene and further contains a XhoI site for insertion of cDNA clones. The functional elements of pMT2 CXM have been described (Kaufman, R.J., 1985, Proc. Natl. Acad. Sci. USA 82:689-693) and include the adenovirus VA genes, the SV40 origin of replication including the 72 bp enhancer, the adenovirus major late promoter including a 5' splice site and the majority of the adenovirus tripartite leader sequence present on adenovirus late mRNAs, a 3' splice acceptor site, a DHFR insert, the SV40 early polyadenylation site (SV40), and pBR322 sequences needed for propagation in E. coli.

Plasmid pMT2 CXM is obtained by EcoRI digestion of pMT2-VWF, which has been deposited with the American Type Culture Collection (ATCC), Rockville, MD (USA) under accession number ATCC 67122. EcoRI digestion excises the cDNA insert present in pMT2-VWF, yielding pMT2 in linear form which can be ligated and used to transform E. coli HB 101 or DH-5 to ampicillin resistance. Plasmid pMT2 DNA can be prepared by conventional methods. pMT2 CXM is then constructed using loopout/in mutagenesis [Morinaga, et al., Biotechnology 84: 636 (1984). This removes bases 1075 to 1145 relative to the Hind III site near the SV40 origin of replication and enhancer sequences of pMT2. In addition it inserts the following sequence:

5' PO-CATGGGCAGCTCGAG-3'

at nucleotide 1145. This sequence contains the recognition site for the restriction endonuclease Xho I. A derivative of pMT2CXM, termed pMT23, contains recognition sites for the restriction endonucleases PstI, Eco RI, SalI and XhoI. Plasmid pMT2 CXM and pMT23 DNA may be prepared by conventional methods.

pEMC2β1 derived from pMT21 may also be suitable in practice of the invention. pMT21 is derived from pMT2 which is derived from pMT2-VWF. As described above EcoRI digestion excises the cDNA insert present in pMT-VWF, yielding pMT2 in linear form which can be ligated and used to transform E. Coli HB 101 or DH-5 to ampicillin resistance. Plasmid pMT2 DNA can be prepared by conventional methods.

pMT21 is derived from pMT2 through the following two modifications. First, 76 bp of the 5' untranslated region of the DHFR cDNA including a stretch of 19 G residues from G/C tailing for cDNA cloning is deleted. In this process, a XhoI site is inserted to obtain the following sequence immediately upstream from DHFR: 5' - CTGCAGGCGAGCCTGAATTCCTCGAGCCATCATG-3'

PstI Eco RI XhoI

Second, a unique ClaI site is introduced by digestion with EcoRV and XbaI, treatment with Klenow fragment of DNA polymerase I, and ligation to a ClaI linker (CATCGATG). This deletes a 250 bp segment from the adenovirus associated RNA (VAI) region but does not interfere with VAI RNA gene expression or function. pMT21 is digested with EcoRI and XhoI, and used to derive the vector pEMC2B1.

A portion of the EMCV leader is obtained from pMT2-ECAT1 [S.K. Jung, et al, J. Virol 63:1651-1660 (1989)] by digestion with Eco RI and PstI, resulting in a 2752 bp fragment. This fragment is digested with TaqI yielding an Eco RI-TaqI fragment of 508 bp which is purified by electrophoresis on low melting agarose gel. A 68 bp adapter and its complementary strand are synthesized with a 5' TaqI protruding end and a 3' XhoI protruding end which has the following sequence:

5'CGAGGTTAAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTC
CTTT
TaqI

GAAAAACACGATTGC-3'
XhoI

This sequence matches the EMC virus leader sequence from nucleotide 763 to 827. It also changes the ATG at position 10 within the EMC virus leader to an ATT and is followed by a XhoI site. A three way ligation of the pMT21 Eco RI-XhoI fragment, the EMC virus EcoRI-TaqI fragment, and the 68 bp
oligonucleotide adapter TaqI-XhoI adapter resulting in the vector pEMC2β1.
This vector contains the SV40 origin of replication and enhancer, the adenovirus major late promoter, a cDNA copy of the majority of the adenovirus tripartite leader sequence, a small hybrid intervening sequence, an SV40 polyadenylation signal and the adenovirus VA I gene, DHFR and β-lactamase markers and an EMC sequence, in appropriate relationships to direct the high level expression of the desired cDNA in mammalian cells.

The construction of vectors may involve modification of the mammalian *cerberus* DNA sequences. For instance, mammalian *cerberus* cDNA can be modified by removing the non-coding nucleotides on the 5' and 3' ends of the coding region. The deleted non-coding nucleotides may or may not be replaced by other sequences known to be beneficial for expression. These vectors are transformed into appropriate host cells for expression of mammalian *cerberus* proteins. Additionally, the sequence of SEQ ID NO:1 other sequences encoding mammalian *cerberus* proteins can be manipulated to express a mature mammalian *cerberus* protein by deleting mammalian *cerberus* encoding

propeptide sequences and replacing them with sequences encoding the complete propeptides of other proteins.

One skilled in the art can manipulate the sequences of SEQ ID NO:1 or SEQ ID NO:7 by eliminating or replacing the mammalian regulatory sequences flanking the coding sequence with bacterial sequences to create bacterial vectors for intracellular or extracellular expression by bacterial cells. For example, the coding sequences could be further manipulated (e.g. ligated to other known linkers or modified by deleting non-coding sequences therefrom or altering nucleotides therein by other known techniques). The modified mammalian *cerberus* coding sequence could then be inserted into a known bacterial vector using procedures such as described in T. Taniguchi et al., Proc. Natl Acad. Sci. USA, 77:5230-5233 (1980). This exemplary bacterial vector could then be transformed into bacterial host cells and a protein expressed thereby. For a strategy for producing extracellular expression of mammalian *cerberus* proteins in bacterial cells, see, e.g. European patent application EPA 177,343.

Similar manipulations can be performed for the construction of an insect vector [See, e.g. procedures described in published European patent application 155,476] for expression in insect cells. A yeast vector could also be constructed employing yeast regulatory sequences for intracellular or extracellular expression of the factors of the present invention by yeast cells. [See, e.g., procedures described in published PCT application WO86/00639 and European patent application EPA 123,289].

A method for producing high levels of a mammalian *cerberus* protein of the invention in mammalian cells may involve the construction of cells containing multiple copies of the heterologous mammalian *cerberus* gene. The heterologous gene is linked to an amplifiable marker, e.g. the dihydrofolate reductase (DHFR) gene for which cells containing increased gene copies can be selected for propagation in increasing concentrations of methotrexate (MTX) according to the procedures of Kaufman and Sharp, J. Mol. Biol., 159:601-629 (1982). This approach can be employed with a number of different cell types.

For example, a plasmid containing a DNA sequence for a mammalian *cerberus* protein of the invention in operative association with other plasmid sequences enabling expression thereof and the DHFR expression plasmid pAdA26SV(A)3 [Kaufman and

Sharp, Mol. Cell. Biol., 2:1304 (1982)] can be co-introduced into DHFR-deficient CHO cells, DUKX-BII, by various methods including calcium phosphate coprecipitation and transfection, electroporation or protoplast fusion. DHFR expressing transformants are selected for growth in alpha media with dialyzed fetal calf serum, and subsequently selected for amplification by growth in increasing concentrations of MTX (e.g. sequential steps in 0.02, 0.2, 1.0 and 5uM MTX) as described in Kaufman et al., Mol Cell Biol., 5:1750 (1983). Transformants are cloned, and biologically active mammalian *cerberus* expression is monitored by assay in one of the assays described in the examples and specification. Mammalian *cerberus* protein expression should increase with increasing levels of MTX resistance. Mammalian *cerberus* polypeptides are characterized using standard techniques known in the art such as pulse labeling with [35S] methionine or cysteine and polyacrylamide gel electrophoresis. Similar procedures can be followed to produce other related *cerberus* proteins.

A mammalian *cerberus* protein of the present invention, which demonstrates *cerberus* activity, has application in the induction, formation, growth, differentiation, proliferation and/or maintenance and healing of cells and tissues such as neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, and other tissues, in humans and other animals. Such a preparation employing mammalian *cerberus* protein may have prophylactic use in treatment of Parkinson's disease, Alzheimer's disease, as well as preventing neural tumors, and other neural tissue disorders. De novo formation of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, and other cells of neural phenotype, induced by a *cerberus* protein contributes to the repair of congenital, trauma induced, or oncologic tissue defects or conditions. Mammalian *cerberus* protein may also be used in the treatment of neural disease, and in other tissue and organ repair processes. Such agents may provide an environment to attract suitable stem cells, stimulate growth and proliferation of neuron-forming cells or induce differentiation of progenitors of neuron-forming cells, and may also support the regeneration of other tissues and organs. Mammalian *cerberus* polypeptides of the invention may also be useful in the treatment of organ disorders.

The proteins of the invention may also be used in wound healing and in related tissue repair. The types of wounds include, but are not limited to burns, incisions and ulcers. (See, e.g. PCT Publication WO84/01106 for discussion of wound healing and related tissue repair). It is further contemplated that proteins of the invention may increase neuronal, astrocytic and/or glial cell survival and therefore be useful in transplantation and treatment of conditions exhibiting a decrease in neuronal survival and repair. The proteins of the invention may further be useful for the treatment of conditions related to other types of tissue, such as epidermis, muscle, connective tissue, such as bone, cartilage, tendon and ligament, and other organs such as pancreas, liver, spleen, lung, cardiac, brain and kidney tissue. The proteins of the present invention may also have value as a dietary supplement, or as additives for cell culture media. For this use, the proteins may be used in intact form, or may be predigested to provide a more readily absorbed supplement.

The proteins of the invention may also have other useful properties characteristic of the *cerberus* family of proteins. Such properties include angiogenic, chemotactic and/or chemoattractant properties, and effects on cells including differentiation responses, cell proliferative responses and responses involving cell adhesion, migration and extracellular matrices. These properties make the proteins of the invention potential agents for wound healing, reduction of fibrosis and reduction of scar tissue formation. The proteins of the invention may also be useful for the induction of formation of cells capable of secreting valuable hormones, including endocrine or exocrine hormones.

A further aspect of the invention is a therapeutic method and composition for treating disorders of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes, and other conditions related to neuronal and neural tissue disorders or diseases. The invention further comprises therapeutic methods and compositions for wound healing and tissue repair. Such compositions comprise a therapeutically effective amount of at least one mammalian *cerberus* protein of the present invention in admixture with a pharmaceutically acceptable vehicle, carrier or matrix. It is further contemplated that compositions of the invention may increase neuronal, glial cell and astrocyte survival and therefore be useful in transplantation and treatment of conditions exhibiting a decrease in neuronal survival.

It is expected that *cerberus* and *cerberus-like* proteins may exist in nature as homodimers or heterodimers. To promote the formation of dimers of *cerberus-like protein* with increased stability, one can genetically engineer the DNA sequence of SEQ ID NO:1 or SEQ ID NO:7 to provide one or more additional cysteine residues to increase potential dimer formation. The resulting DNA sequence would be capable of producing a "cysteine added variant" of *cerberus-like* protein. Alternatively, one can produce "cysteine added variants" of *cerberus-like* proteins by altering the sequence of the protein at the amino acid level, for example, by altering the amino acid sequences of one or more amino acid residues to Cys. Production of "cysteine added variants" of proteins is described in United States Patent 5,166,322, the disclosure of which is hereby incorporated by reference.

It is expected that the proteins of the invention may act in concert with or perhaps synergistically with other related proteins and growth factors. Such combinations may comprise separate molecules of the *cerberus* or *cerberus-like* proteins and other proteins or heteromolecules comprised of different moieties. For example, a method and composition of the invention may comprise a disulfide linked dimer comprising a *cerberus* protein subunit and a subunit from one of the "BMP" proteins. Thus, the present invention includes a purified *cerberus-like* polypeptide which is a heterodimer wherein one subunit comprises an amino acid sequence of SEQ ID NO:2 or SEQ ID NO:8, and one subunit comprises an amino acid sequence for a bone morphogenetic protein selected from the group consisting of BMP-2, BMP-3, BMP-4, BMP-5, BMP-6, BMP-7, BMP-8, BMP-9, BMP-10, BMP-11, BMP-12 or BMP-13, disclosed in PCT application WO 95/16035, or BMP-15, disclosed in PCT application WO96/36710 or BMP-16, disclosed in co-pending patent application serial number 08/715/202, filed September 18, 1996. A further embodiment may comprise a heterodimer of *cerberus* moieties, for example, of *Xenopus cerberus* and a mammalian homologue of *Xenopus cerberus* or other *cerberus-like* protein. Further therapeutic methods and compositions of the invention therefore comprise a therapeutic amount of at least one mammalian *cerberus* protein of the invention with a therapeutic amount of at least one other protein, such as a member of the TGF-β superfamily of proteins, which includes the bone morphogenetic proteins (BMPs), growth and differentiation factors (GDFs) and other

proteins. The composition may include other agents and growth factors such as epidermal growth factor (EGF), fibroblast growth factor (FGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), *Wnts*, *hedgeghogs*, including sonic, indian and desert *hedgehogs*, activins, inhibins, and k-fibroblast growth factor (kFGF), parathyroid hormone (PTH), leukemia inhibitory factor (LIF/HILDA/DIA), insulin-like growth factors (IGF-I and IGF-II). Portions of these agents may also be used in compositions of the present invention.

The preparation and formulation of such physiologically acceptable protein compositions, having due regard to pH, isotonicity, stability and the like, is within the skill of the art. The therapeutic compositions are also presently valuable for veterinary applications due to the lack of species specificity in *cerberus* proteins. Particularly domestic animals and thoroughbred horses in addition to humans are desired patients for such treatment with the *cerberus* proteins of the present invention.

The therapeutic method includes administering the composition topically, systemically, or locally as by injection or implantation. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes or other tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than the *cerberus* proteins which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the *cerberus* composition in the methods of the invention.

For implantation, the composition preferably includes a matrix capable of delivering mammalian *cerberus* proteins to the site of neurons and/or related neural cells and tissues such as Schwann cells, glial cells and astrocytes or other tissue damage, providing a structure for the developing tissue and optimally capable of being resorbed into the body. The matrix may provide slow release of mammalian *cerberus* and/or other protein, as well as proper presentation and appropriate environment for cellular infiltration. Such matrices may be formed of materials presently in use for other

implanted medical applications. The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the mammalian *cerberus* compositions will define the appropriate formulation.

The dosage regimen will be determined by the attending physician considering various factors which modify the action of the mammalian *cerberus* protein, e.g. amount of tissue desired to be formed, the site of tissue damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue, the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and the types of mammalian *cerberus* proteins in the composition. The addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage.

Progress can be monitored by periodic assessment of tissue growth and/or repair. The progress can be monitored, for example, x-rays, histomorphometric determinations and tetracycline labeling.

The following examples illustrate practice of the present invention in recovering and characterizing mammalian *cerberus* protein and employing the DNA to recover other *cerberus* proteins, obtaining the human proteins and expressing the proteins via recombinant techniques. As will be recognized, numerous variations of the materials and methods described can be prepared and are within the invention.

**Example 1: Cloning of a Murine Homologue to *Xenopus cerberus***

The carboxy-terminal, cysteine rich domain of xenopus *cerberus* was used to screen mammalian libraries and ESTS reported in GENBANK. An EST, AA120122, derived from the Beddington day 7.5 embryonic region library was identified by sequence homology to xenopus *cerberus*. While the partial clone has almost no sequence conservation in the N-terminal half of the molecule, the C-terminal domain and especially the pattern of the nine cysteines showed sequence conservation. The full-length cDNA encoding the murine *cerberus*-like protein was isolated from a murine embryonic cell cDNA library and the gene isolated from a murine genomic library.

Murine *cerberus-like* cDNA (SEQ ID NO:1) encodes a 272 amino acid protein with predicted MW 30.5 Kd (SEQUENCE ID NO:2). The murine *cerberus-like* gene, Mcerb-1, has a single 2Kb intron at position 564 of the cDNA SEQ ID NO:1. The predicted protein contains a hydrophobic signal sequence at its amino terminus, indicating the molecule is secreted. Analysis by Sigcleave indicates that the first 17 resides are cleaved from the mature molecule upon secretion (Sigcleave score = 7.6). The mature murine *cerberus-like* protein is predicted to be a 255 residue protein of MW 28.6 Kd including residues 18-272 of SEQ ID NO:2. The *cerberus-like* cDNA was expressed in COS cells and labeled with $^{35}$S-Met/$^{35}$S-Cys. The resulting protein was secreted into the culture medium and resulted in a smeared band of MW 38-44 Kd when fractionated on 16% polyacrylamide reducing gels. The smeared bands indicate that the protein is glycosylated consistent with the two putative N-linked glycosylation sites in the sequence. Expression of the protein in reticulocyte systems shows a 33kD band consistent with predicted protein in the absence of glycosylation. When fractionated on non-reducing polyacrylamide gels, approximately half of the COS expressed material runs with a molecular weight of 78-84 Kd indicating the protein is capable of forming cysteine-linked homodimers. The protein sequence contains a hydrophobic signal sequence at its amino terminus and a cysteine-rich domain close to its carboxy terminus. The cysteine-rich domains of *Xenopus cerberus* and murine *cerberus-like* proteins are 58% identical and the overall amino acids are 31% identical. The nine cysteine pattern at the carboxy terminus is conserved between *Xenopus cerberus* and mammalian *cerberus-like* protein.

**Example 2: Mammalian *cerberus-like* Encodes a Secreted Protein.**

To test whether the full length cDNA encoded a secreted protein, the 293T human cell line was transiently transfected with the DNA sequence of SEQ ID NO:1, which encodes mammalian *cerberus-like* protein, cloned in an eukaryotic expression vector and labeled with $^{35}$S-methionine. A broad band was secreted into the culture medium, as well as a minor form. Similar to the COS cell expression, fractionation of the material expressed in 293T cells evidenced the formation of cysteine-linked dimers. The *cerberus-like* protein is presumably glycosylated since the protein translated in the reticulocyte system (in the absence of membranes) results in a band of 33 kD molecular

weight corresponding to the molecular weight predicted from the amino acid sequence. *Cerberus-like* protein has two putative N-linked glycosylation sites. Accordingly, the DNA of SEQ ID NO:1 defines a secreted protein with high amino acid identity to *Xenopus cerberus* in the cysteine-rich domain.

**Example 3: Expression of Mammalian *Cerberus-like* Protein**

In prestreak mouse embryos (5.5 days post coitum), mammalian *cerberus-like* transcripts were detected on one side of the primitive endoderm, including the distal tip of the embryo. At early primitive streak, expression was found in a patch of primitive endoderm cells on one side of the embryonic region and no longer extends to the tip of the embryo. This patch corresponded to the anterior side of the embryo, because in sections it is found in the endoderm opposite to the forming primitive streak which can be recognized as a thickening of the posterior epiblast. At mid-streak, the *cerberus-like* positive area remained in the anterior primitive endoderm. At the late streak stage, when the node has reached the distal tip of the embryo, a second population of *cerberus-like* expressing was seen in the region surrounding the node. These cells presumably correspond to definitive endodermal cells, that are known to emerge from the node. The endodermal nature of this cell population was confirmed by histological analysis. At the neural plate stage, *cerberus-like* is found underlying the anterior neural plate, in a pattern comparable to the domain of *Otx2* expression in endoderm. Ang et al., Cell, 78:561-574 (1994). At this stage, the *cerberus-like* positive cell population presumably consists of both primitive and definitive endoderm, and do not include the node itself. Importantly, cells expressing the *cerberus-like* secreted factor are in direct contact with cells that subsequently give rise to the fore- and midbrain region of the CNS.

At the early headfold stage, *cerberus-like* signal in the endoderm stats to weaken. At late headfold, *cerberus-like* expression is confined to the midline and adjoining endoderm. Expression is found in all cells of the midline from the rostral end of the embryo to the proximity of the node and includes anterior gut endoderm and mesoderm from the prechordal and notochordal plates. Expression of *cerberus-like* mRNA in anterior endoderm remains until the start of somitogenesis and then becomes undetectable. Concomitantly, a late phase of expression begins in the mesoderm of the somites. Thus, *cerberus-like* defines an anterior domain of the endoderm in early mouse

embryos. Expression is found in anterior endoderm in direct contact with the future neural plate, but never detected in posterior endoderm cells in contact with the primitive streak, providing further evidence in support of the role of *cerberus-like* in the induction of anterior neuroectoderm.

**Example 4: Assays to Determine Function of *Cerberus* or *Cerberus-Like* in the Mouse**

To determine function of *cerberus* and *cerberus-like* genes in the mouse embryo, transgenic knockout and misexpression mice can be made by engineering of murine embryonic stem (ES) cells and injection into the blastula using standard procedures. The murine knockout can be achieved by replacing a central portion of the cerberus or cerberus like genes with a selectable marker (*e.g., neo*), transfecting the construct into ES cells and selecting for the double crossover. For misexpression, the *cerberus* and *cerberus-like* genes can be engineered to be expressed from generic (*e.g.*, actin) or tissue specific (*e.g.*, IDX) promoters and reintroduced into murine ES cells. Transgenic animals can be generated from such engineered cells using classical procedures.

To determine the function of the cerberus or cerberus like proteins in the adult mouse, the protein can be directly injected into tissues or delivered by viral vectors. For example, the *cerberus* or *cerberus-like* genes can be transiently overexpressed in the adult mouse using adenovirus vectors and the function or activity of the proteins investigated by physiological, histochemical and biochemical analysis of the animals.

**Example 5: Isolation of Human *Cerberus* and *Cerberus-like* Genes.**

Hybridization of the murine *cerberus-like* gene to human DNA sequences identified a single *Xba*1 generated band of approximately 15 Kb demonstrating its utility as probe to identify the human *cerberus-like* gene within a human genomic library. Alignment of the *Xenopus cerberus* and murine *cerberus-like* genes has identified three regions of sufficient nucleic acid homology within the cysteine rich domain to serve as probes for the human cerberus and *cerberus-like* genes. The consensus sequences for the three probes are:

(1) TGCCCTTCAMYCAGAMYATTGYMCATGAARACTGT [SEQ ID NO:3];

(2) CAGAACAAYCTKTGCTTTGGTAAATGCA [SEQ ID NO:4]; and

(3) TGYTCCCAYTGCTYGCCSWCCAAATT [SEQ ID NO: 5].

The first probe spans the 2 Kb intron in the mouse gene and is therefore less likely to function well as a probe. Used individually or together, the latter two oligos containing 6-fold and 10-fold degeneracy, respectively, would serve as probes for human *cerberus* and *cerberus-like* genes in genomic or cDNA libraries.

The nucleotide and amino acid sequences obtained for human *cerberus* according to the procedures described herein are shown in SEQ ID NO:7 and SEQ ID NO:8, respectively.

**Example 6: Identification and Isolation of Additional *Cerberus* and *Cerberus-like* Proteins.**

The *cerberus* and *cerberus-like* proteins are members of a family which can be recognized by their unique cysteine pattern. This family includes *cerberus*, *cerberus-like* protein, *Dan* protein and *Norrie* protein. *Dan* protein is a tumor suppressor candidate and defects in the Norrie protein result in congenital defects including blindness and deafness. Members of the *cerberus* family thus seem to play an important role in cell differentiation and proliferation and thus it is important to find other members of this family of proteins. Family members cannot necessarily be recognized by amino acid homology, which demonstrates a significant amount of diversity, but can be recognized by a unique pattern of nine cysteines. The consensus cysteine pattern in each of these proteins is: C ($X_{13-15}$) C ($X_9$) CxGxC ($X_{14-23}$) CXXC ($X_{13}$) C ($X_{15-18}$) CXC, wherein $X_y$ indicates the number of non-cysteine amino acid residues found between each cysteine in the conserved cysteine pattern of the *cerberus* and *cerberus-like* families of proteins.

By searching mammalian ESTs for this motif, novel members of the cerberus family have been identified including the human EST N35377 and the murine EST AA289245. Using standard procedures, the full length genes can be isolated from either genomic or cDNA libraries. These genes are expected to encode signaling proteins which function to pattern the embryo, control cellular differentiation or cell proliferation and thus be candidate proteins in medical therapy.

Within the cysteine-rich domains of *Xenopus cerberus* and murine *cerberus-like* proteins are two discrete subregions which are highly conserved cysteine motifs and may be particularly useful for the identification and isolation of *cerberus* and *cerberus-like* proteins from other species, as well as related members of the *cerberus* and *cerberus-like*

protein families. Use of these highly conserved motifs to screen mammalian libraries and ESTS reported in GENBANK is expected to identify additional proteins of the *cerberus* and *cerberus-like* families of proteins from humans and other species, as well as identify further family members may be found. The first region is the motif C-X-G-X-C, which corresponds to Cys-Phe-Gly-Lys-Cys found at amino acid residues 186 to 190 of SEQ ID NO:1. Thus, degenerate oligonucleotides to the sequence TGC TTT GGC AAA TGC at nucleotide positions 613 to 627 of SEQ ID NO:1, and the adjoining regions, may be useful to identify and isolate other genes which share the C-X-G-X-C motif of *cerberus* and *cerberus-like* proteins. A second highly conserved region within the cysteine-rich domains of *Xenopus cerberus* and murine *cerberus-like* proteins is the motif C-X-X-C, which corresponds to Cys-Ser-His-Cys found at amino acid residues 206 to 209 of SEQ ID NO:1. Thus, degenerate oligonucleotides to the sequence TGC TTC CAC TGC at nucleotide positions 673 to 684 of SEQ ID NO:1, and the adjoining regions, may be useful to identify and isolate other genes which share the C-X-X-C motif of *cerberus* and *cerberus-like* proteins. Using the above motifs, additional proteins of the *cerberus* and *cerberus-like* families of proteins may be identified from other species, and further family members from humans, mice or frogs may be found. Use of the two motifs above in concert may further provide specific identification and isolation of additional *cerberus* and *cerberus-like* family members.

**Example 7: Microinjection Assays for *Cerberus* and *Cerberus-Like* Activity**

Injection of mRNA encoding *xenopus cerberus* into *xenopus* blastomeres at different stages (4 cell, 8 cell and 32 cell) has profound effects on the expression pattern of specific genes and on the resulting morphology of the developing embryo. Misexpression of *cerberus* inhibits development of the prechordal plate, notochord and ventral trunk mesoderm and their respective molecular markers, goosecoid, collagen II and x-globin. Microinjection of *cerberus* induces anterior neuroectodermal structures such as brain, olfactory placodes and cement gland. *Cerberus* induced gene expression includes N-CAM (brain), *Otx2* (anterior brain), *CG-13* (cement gland) and *Nkx-2.5* (heart primordium). Induction of neural tissue by *cerberus* was specific to anterior region of the brain as indicated by upregulation of the *Otx2* marker but not more posterior markers including *En-2* (midbrain-hind brain junction), *Krox-20* (hind brain) and *X1Hbox-6*

(spinal cord). Injection of *cerberus* into specific cells of the 32-cell blastomere resulted in the induction of ectopic heads, and duplicate heart and liver.

Like *cerberus*, microinjection of *cerberus-like* mRNA into animal cap explants induces anterior CNS in *xenopus* embryos. However, microinjection of mammalian *cerberus-like* mRNA into *xenopus* embryos did not induce formation of ectopic head structures, for example, containing forebrain, cyclopic eyes, olfactory placodes and cement glands, suggesting overlapping but not identical functional effects. Thus, *cerberus-like* is a neuralizing factor, that leads to the formation of forebrain in *xenopus* assays.

**Example 8. Additional Embryonic Stem Cell Assay**

In order to assay the effects of the *cerberus-like* proteins of the present invention, it is possible to assay the growth and differentiation effects *in vitro* on a number of available embryonic stem cell lines. One such cell line is ES-E14TG2, which is available from the American Type Culture Collection in Rockville, Md.

In order to conduct the assay, cells may be propagated in the presence of 100 units of LIF to keep them in an undifferentiated state. Assays are setup by first removing the LIF and aggregating the cells in suspension, in what is known as embryoid bodies. After 3 days the embryoid bodies are plated on gelatin coated plates (12 well plates for PCR analysis, 24 well plates for immunocytochemistry) and treated with the proteins to be assayed. Cells are supplied with nutrients and treated with the protein factor every 2-3 days. Cells may be adapted so that assays may be conducted in media supplemented with 15% Fetal Bovine Serum (FBS) or with CDM defined media containing much lower amounts of FBS.

At the end of the treatment period (ranging from 7-21 days) RNA is harvested from the cells and analyzed by quantitative multiplex PCR for the following markers: *Brachyury*, a mesodermal marker, *AP-2*, an ectodermal marker, and *HNF-3α* an endodermal marker. Through immunocytochemistry, it is also possible to detect the differentiation of neuronal cells (glia, astrocytes and neurons), muscle cells (cardiomyocytes, skeletal and smooth muscle), and various other phenotype markers such as proteoglycan core protein (cartilage), and cytokeratins (epidermis). Since these cells

have a tendency to differentiate autonomously when LIF is removed. the results are always quantitated by comparison to an untreated control.

**Example 9 - Expression of Human *Cerberus* Protein**

The human *Cerberus* cDNA sequence (SEQ ID NO:7) has been expressed in COS cells using mammalian expression vectors as described in the application. The observed protein was secreted, with a molecular weight of approximately 35-45 kD, consistent with a glycosolated protein of 250 (amino acids #18 to #267 of SEQ ID NO:8) amino acids in length.

The foregoing descriptions detail presently preferred embodiments of the present invention. Numerous modifications and variations in practice thereof are expected to occur to those skilled in the art upon consideration of these descriptions. Those modifications and variations are believed to be encompassed within the claims appended hereto.

All of the publications and patents referred to herein are hereby specifically incorporated by reference, as if fully set forth herein for the referred to disclosure.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(I) APPLICANT: FOLLETTIE, MAXIMILLIAN
DeROBERTIS, EDWARD M.

(ii) TITLE OF INVENTION: Mammalian Cerberus-Like Protein & Compositions

(iii) NUMBER OF SEQUENCES: 8

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Genetics Institute, Inc.
(B) STREET: 87 CambridgePark Drive
(C) CITY: Cambridge
(D) STATE: Massachusetts
(E) COUNTRY: US
(F) ZIP: 02140

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: TBD
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: LAZAR, STEVEN R
(B) REGISTRATION NUMBER: 32,618
(C) REFERENCE/DOCKET NUMBER: GI 5290APCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617) 498-8260
(B) TELEFAX: (617) 876-5851

(2) INFORMATION FOR SEQ ID NO:1:

(I) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1003 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 58..873

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCGGCC AAAGAGGCCT ATGTGAATCT AACCTCAGTC TCTGGGAATC AGGAAGC 57

ATG CAT CTC CTC TTA GTT CAG CTG CTT GTT CTC TTG CCT CTG GGG AAG 105

```
Met His Leu Leu Leu Val Gln Leu Leu Val Leu Leu Pro Leu Gly Lys
  1               5                  10                  15

GCA GAC CTA TGT GTG GAT GGC TGC CAG AGT CAG GGC TCT TTA TCC TTT      153
Ala Asp Leu Cys Val Asp Gly Cys Gln Ser Gln Gly Ser Leu Ser Phe
            20                  25                  30

CCT CTC CTA GAA AGG GGT CGC AGA GAT CTC CAC GTG GCC AAC CAC GAG      201
Pro Leu Leu Glu Arg Gly Arg Arg Asp Leu His Val Ala Asn His Glu
        35                  40                  45

GAG GCA GAA GAC AAG CCG GAT CTG TTT GTG GCC GTG CCA CAC CTC ATG      249
Glu Ala Glu Asp Lys Pro Asp Leu Phe Val Ala Val Pro His Leu Met
    50                  55                  60

GGC ACC AGC CTG GCT GGG GAA GGC CAG AGG CAG AGA GGG AAG ATG CTG      297
Gly Thr Ser Leu Ala Gly Glu Gly Gln Arg Gln Arg Gly Lys Met Leu
65                  70                  75                  80

TCC AGG CTT GGA AGA TTC TGG AAG AAA CCT GAG ACC GAA TTT TAC CCC      345
Ser Arg Leu Gly Arg Phe Trp Lys Lys Pro Glu Thr Glu Phe Tyr Pro
                85                  90                  95

CCA AGG GAT GTG GAA AGC GAT CAT GTC TCA TCG GGG ATG CAG GCC GTG      393
Pro Arg Asp Val Glu Ser Asp His Val Ser Ser Gly Met Gln Ala Val
            100                 105                 110

ACT CAG CCA GCA GAT GGG AGG AAA GTG GAG AGA TCA CCT CTA CAG GAG      441
Thr Gln Pro Ala Asp Gly Arg Lys Val Glu Arg Ser Pro Leu Gln Glu
        115                 120                 125

GAA GCC AAG AGG TTC TGG CAT CGG TTC ATG TTC AGA AAG GGC CCG GCG      489
Glu Ala Lys Arg Phe Trp His Arg Phe Met Phe Arg Lys Gly Pro Ala
    130                 135                 140

TTC CAG GGA GTC ATC CTG CCC ATC AAA AGC CAC GAA GTA CAC TGG GAG      537
Phe Gln Gly Val Ile Leu Pro Ile Lys Ser His Glu Val His Trp Glu
145                 150                 155                 160

ACC TGC AGG ACT GTG CCC TTC AAC CAG ACC ATT GCC CAT GAA GAC TGT      585
Thr Cys Arg Thr Val Pro Phe Asn Gln Thr Ile Ala His Glu Asp Cys
                165                 170                 175

CAA AAA GTC GTT GTC CAG AAC AAC CTT TGC TTT GGC AAA TGC AGT TCC      633
Gln Lys Val Val Val Gln Asn Asn Leu Cys Phe Gly Lys Cys Ser Ser
            180                 185                 190

ATT CGT TTT CCC GGA GAA GGG GCA GAT GCC CAC AGC TTC TGC TCC CAC      681
Ile Arg Phe Pro Gly Glu Gly Ala Asp Ala His Ser Phe Cys Ser His
        195                 200                 205

TGC TCG CCC ACC AAA TTC ACC ACC GTG CAC TTG AGG CTG AAC TGC ACC      729
Cys Ser Pro Thr Lys Phe Thr Thr Val His Leu Arg Leu Asn Cys Thr
    210                 215                 220

AGC CCA ACC CCC GTG GTC AAG ATG GTG ATG CAA GTA GAA GAG TGT CAG      777
Ser Pro Thr Pro Val Val Lys Met Val Met Gln Val Glu Glu Cys Gln
225                 230                 235                 240

TGC ATG GTG AAG ACG GAA CGT GGA GAG GAG CGC CTC CTA CTG GCT GGT      825
Cys Met Val Lys Thr Glu Arg Gly Glu Glu Arg Leu Leu Leu Ala Gly
                245                 250                 255
```

```
TCC CAG GGT TCC TTC ATC CCT GGA CTT CCA GCT TCA AAA ACA AAC CCA      873
Ser Gln Gly Ser Phe Ile Pro Gly Leu Pro Ala Ser Lys Thr Asn Pro
            260                 265                 270

TGATTACCTC AACAGAAAGC AAAACCTCAA CAGAATAAGT GAGGGTTATT CAATCTGGAA   933

ATGTTATGTG AGTTATATAA AGATCAGTGG AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA   993

AAGCGGCCGC                                                         1003
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 272 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met His Leu Leu Leu Val Gln Leu Leu Val Leu Leu Pro Leu Gly Lys
 1               5                  10                  15

Ala Asp Leu Cys Val Asp Gly Cys Gln Ser Gln Gly Ser Leu Ser Phe
            20                  25                  30

Pro Leu Leu Glu Arg Gly Arg Arg Asp Leu His Val Ala Asn His Glu
        35                  40                  45

Glu Ala Glu Asp Lys Pro Asp Leu Phe Val Ala Val Pro His Leu Met
    50                  55                  60

Gly Thr Ser Leu Ala Gly Glu Gly Gln Arg Gln Arg Gly Lys Met Leu
65                  70                  75                  80

Ser Arg Leu Gly Arg Phe Trp Lys Lys Pro Glu Thr Glu Phe Tyr Pro
                85                  90                  95

Pro Arg Asp Val Glu Ser Asp His Val Ser Ser Gly Met Gln Ala Val
            100                 105                 110

Thr Gln Pro Ala Asp Gly Arg Lys Val Glu Arg Ser Pro Leu Gln Glu
        115                 120                 125

Glu Ala Lys Arg Phe Trp His Arg Phe Met Phe Arg Lys Gly Pro Ala
    130                 135                 140

Phe Gln Gly Val Ile Leu Pro Ile Lys Ser His Glu Val His Trp Glu
145                 150                 155                 160

Thr Cys Arg Thr Val Pro Phe Asn Gln Thr Ile Ala His Glu Asp Cys
                165                 170                 175

Gln Lys Val Val Val Gln Asn Asn Leu Cys Phe Gly Lys Cys Ser Ser
            180                 185                 190

Ile Arg Phe Pro Gly Glu Gly Ala Asp Ala His Ser Phe Cys Ser His
        195                 200                 205

Cys Ser Pro Thr Lys Phe Thr Thr Val His Leu Arg Leu Asn Cys Thr
    210                 215                 220
```

Ser Pro Thr Pro Val Val Lys Met Val Met Gln Val Glu Glu Cys Gln
225 230 235 240

Cys Met Val Lys Thr Glu Arg Gly Glu Glu Arg Leu Leu Leu Ala Gly
245 250 255

Ser Gln Gly Ser Phe Ile Pro Gly Leu Pro Ala Ser Lys Thr Asn Pro
260 265 270

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TGCCCTTCAM YCAGAMYATT GYMCATGAAR ACTGT 35

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CAGAACAAYC TKTGCTTTGG TAAATGCA 28

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TGYTCCCAYT GCTYGCCSWC CAAATT 26

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3595 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 117..623

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 624..2402

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2403..2699

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TCCGGCCAGG CAGGTATCTA TATATCCGAT TTCCTTTTTC CCAAGTCCTG CAGAAGAATG 60

AGCCTCTCCT TTGGGCCTCA TCATTTTACC AAAAAGAAGC TTGGGCCCCT GACAGCATGC 120

ATCTCCTCTT ATTTCAGCTG CTGGTACTCC TGCCTCTAGG AAAGACCACA CGGCACCAGG 180

ATGGCCGCCA GAATCAGAGT TCTCTTTCCC CCGTACTCCT GCCAAGGAAT CAAAGAGAGC 240

TTCCCACAGG CAACCATGAG GAAGCTGAGG AGAAGCCAGA TCTGTTTGTC GCAGTGCCAC 300

ACCTTGTAGC CACCAGCCCT GCAGGGGAAG GCCAGAGGCA GAGAGAGAAG ATGCTGTCCA 360

GATTTGGCAG GTTCTGGAAG AAGCCTGAGA GAGAAATGCA TCCATCCAGG GACTCAGATA 420

GTGAGCCCTT CCCACCTGGG ACCCAGTCCC TCATCCAGCC GATAGATGGA ATGAAAATGG 480

AGAAATCTCC TCTTCGGGAA GAAGCCAAGA AATTCTGGCA CCACTTCATG TTCAGAAAAA 540

CTCCGGCTTC TCAGGGGGTC ATCTTGCCCA TCAAAAGCCA TGAAGTACAT TGGGAGACCT 600

GCAGGACAGT GCCCTTCAGC CAGGTATGTG TTCTGGGGGG AGAGCAGGTA AGAGTTTGCA 660

GGTGGTAGTG GACAGCTGGG ATGGATGGAG AGTAGGGGAA AAGGCTGTCA GGAGCCTGAC 720

TCTAGCTTAA CTACAGATTT GGTCCTTGGG CATTCATCAT AGGATTTGGC AAAGATTAAG 780

TTTCCTTCTG GCCTTTCCAT TTTTTCTTGG CATTGTGGAA ATGCTGCAAG AATGATATGA 840

TGATACTGTC AATATCAGTA ATCATTCATT CACACTGAAG ACACAGAGCT CTGTTTTATT 900

TATTTATTTT TGCATTGGAG GTGATCTACT CAGAGATATA AGTCAGACTG TACCCTCAGT 960

TAGGAAACTG AGAATTTAGA GAATCACCAG AACTCCTCTG TAGCTATCTT TCTGCACTCT 1020

ATTAATATGG GATGAGCAGG TCAACTCCCA TTTGTTGATA AAGTGGGGTG CATTGGACTC 1080

CTTCCCAAAT ACTCTCATAT CCATTTACGA TGGTCTTAAT CCCCATAGTC CATACTTAAT 1140

TACTTTATAG GTTTATGAGG GACTTCTTTA ATAGCTTGCT AAAGCTTATC CCACAACCTC 1200

AAAGTACGTT GAGGTTCTCA GGCAAAAGTT GTCATATCAT TTCTAGTATT ATGATAGCAA 1260

AAAAGTGATT TTCTTTCACT TATTTTCTCA TATGAGCTTT TTAAAAAATC AATCTTGATG 1320

TGAGATCATA TCTCCTCCCC TTAGAAGTAC CTTTCTCCTG ATTCATGTTG TGTTGGCTGA 1380

TTTGTAGTTA TTATGATCAA TTCCATGCTA TTAAGACAAA GGGACATCCT ACTGTCTACT 1440

TCCTCTGGCA ATATCTACAT TCCAAATGTT AAATTAAAAT TGAGAACTTG CATTAGGTCC 1500

TTAAACATGA AGATATTGAA CCAAAAACAT GCAGGGTAGA GTAAAATTTT ATAGTCGAGT 1560

AATGCTACCC AATTAAGCAA GCAATAGAAT AGGGCAATTG ACTGTTCAAG GCAGTTAAGT 1620

ATTCTGCCTG AAAAGGCAAG GATATGTAGC AATGGCAAGT CAATTATCAA ATAATAATGA 1680

CTACTCTGTT GGCCATGTGC AATTAGAAAA TTACCCCTAA GAATCAGGCA ATCAAATTTC 1740

TTTTGAAATT CTTCTTTTGA ATTCTATTGC TAATTAAATT AAAACTAAGA TGTTTGACTC 1800

TTACATATTT TGAAAGGCAT ATAAAGCTAG GTGCTTGGAG TTATGAGAGG TAAAGGTGAT 1860

GTAATATNCA ATGATTTGCA GGCATATGCA TTGTAACTCT GCTTGCATAC AACTTCATAG 1920

ACTTGAATGT ACTACAGGTC TTGCAGAATA GGATAGAATT AAACCTAGAA TGTTCTGATC 1980

TATTCTACGA TCAATGTAAC AAATATGTAT TGGGAGCCTA CTATGCACAA AGCCCTGTGA 2040

GGAATAAAAA AGTAAGGCAC ATTACTTATG TAAGATAATT ACCATTAGAA TTTTTCAATC 2100

GCTCACATCC AATTAGACAA AATTGCTTAA GGTTTTGCAC GAATAATGTA GAGTTAAATA 2160

TTTTTTATGT TAACTTAGGG ATTCCCTAAA GGCTGTTTAA TAATTTACTC AATAAAGAAA 2220

ATTTAATTGA GGTGGTTCTG TGCCCTTATA GATACCATCA CTTGCATATT GCAAATTGTA 2280

TCCAAAATTG GAAAGCTTTG AAATTTTTAA ATTATCCTCA GATTTACAGT CCATAGCTTC 2340

TGCATTATGT GTGTTAAAGA AATAATTCAA AATAACGTAA TGGAAATGTG TTTGCTTTTT 2400

AGACTATAAC CCACGAAGGC TGTGAAAAAG TAGTTGTTCA GAACAACCTT TGCTTTGGGA 2460

AATGCGGGTC TGTTCATTTT CCTGGAGCCG CGCAGCACTC CCATACCTCC TGCTCTCACT 2520

GTTTGCCTGC CAAGTTCACC ACGATGCACT TGCCACTGAA CTGCACTGAA CTTTCCTCCG 2580

TGATCAAGGT GGTGATGCTG GTGGAGGAGT GCCAGTGCAA GGTGAAGACG GAGCATGAAG 2640

ATGGACACAT CCTACATGCT GGCTCCCAGG ATTCCTTTAT CCCAGGAGTT TCAGCTTGAA 2700

GAGCTATCCC ACTATTACCT TTGAAAAGCA AAACCACAAC AGCAAAGATG CTGATTATTC 2760

AGTCTGAAAA TGTTAAGTGG GTACATAACA TTTTCAGGGA AAGGTGACTT GAAACGTAGT 2820

TTTAAATTAG AACGATAGAG GAAATGATAT TAGTCTAGTT ATTGGTACAC GTTTGAGACC 2880

TTGTCTCAGC TCTGCCACTA ACTAGCCGTA GAATGTTAAG TTGTAAAACC TTTCTCCATC 2940

TAAAGATTTT CATCTATAAA TGACGGACCC GACCTAGATG ATTGCTAAAA TCCTTTCCAC 3000

TACTAATATT CCGTGATGCA TTTTCTCCAA GTTTGGGTAA AAGCCCTCCA TCTAAAGAAG 3060

GAAAAGAAAT AAGCGAGACC ATAAAAATGG GCTTCTTTAA TGTGTGTCAA ATCACCAGCA 3120

AGCAAAGAAG CAAGATAGAG AGGGAGGAAG GAAGGAAGGA AGGAAGGAAA GAAGGAAGGA 3180

AGGAAGGAAG GAAGGAAGGA AGGAAGGAAG GAAGGAAGGA AGGAAGGCAG GCAGGCAGGC 3240

AGGCAGGCAG GCAGGCACGC AGGCAGGCAG GGAGGCAGGC TACGTGAAAT ATTTGTAGGA 3300

AAGATTCTCA TACTTATAGT TACTTTTGCA ACCCAAACAG TGTTTTACTT GACTTCTATC 3360

TGATGATTAA GTCTTTCCAC AGATGTAAGG AGTAACTTGC TTGGTTGCCT CCTTTTTAAC 3420

AATACTCCTC ATATAAAGTA CTTAATGTCA GGTCTCTGAC TTTGAAGAAG GAACAGTGAT 3480

GTTAATTTTA GTAGTTTATA TAGGGAAGAG GAACAATCAC TGGTAGCCAA ACAAGTACCT 3540

ATATTATGAG GAAGGAAAAA TACATGACTA CTACCAGGTT TAGAGATCCG AATTC 3595

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 804 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..804

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
ATG CAT CTC CTC TTA TTT CAG CTG CTG GTA CTC CTG CCT CTA GGA AAG        48
Met His Leu Leu Leu Phe Gln Leu Leu Val Leu Leu Pro Leu Gly Lys
 1               5                  10                  15

ACC ACA CGG CAC CAG GAT GGC CGC CAG AAT CAG AGT TCT CTT TCC CCC        96
Thr Thr Arg His Gln Asp Gly Arg Gln Asn Gln Ser Ser Leu Ser Pro
            20                  25                  30

GTA CTC CTG CCA AGG AAT CAA AGA GAG CTT CCC ACA GGC AAC CAT GAG       144
Val Leu Leu Pro Arg Asn Gln Arg Glu Leu Pro Thr Gly Asn His Glu
        35                  40                  45

GAA GCT GAG GAG AAG CCA GAT CTG TTT GTC GCA GTG CCA CAC CTT GTA       192
Glu Ala Glu Glu Lys Pro Asp Leu Phe Val Ala Val Pro His Leu Val
    50                  55                  60

GCC ACC AGC CCT GCA GGG GAA GGC CAG AGG CAG AGA GAG AAG ATG CTG       240
Ala Thr Ser Pro Ala Gly Glu Gly Gln Arg Gln Arg Glu Lys Met Leu
65                  70                  75                  80

TCC AGA TTT GGC AGG TTC TGG AAG AAG CCT GAG AGA GAA ATG CAT CCA       288
Ser Arg Phe Gly Arg Phe Trp Lys Lys Pro Glu Arg Glu Met His Pro
                85                  90                  95

TCC AGG GAC TCA GAT AGT GAG CCC TTC CCA CCT GGG ACC CAG TCC CTC       336
Ser Arg Asp Ser Asp Ser Glu Pro Phe Pro Pro Gly Thr Gln Ser Leu
            100                 105                 110
```

```
ATC CAG CCG ATA GAT GGA ATG AAA ATG GAG AAA TCT CCT CTT CGG GAA      384
Ile Gln Pro Ile Asp Gly Met Lys Met Glu Lys Ser Pro Leu Arg Glu
        115                 120                 125

GAA GCC AAG AAA TTC TGG CAC CAC TTC ATG TTC AGA AAA ACT CCG GCT      432
Glu Ala Lys Lys Phe Trp His His Phe Met Phe Arg Lys Thr Pro Ala
    130                 135                 140

TCT CAG GGG GTC ATC TTG CCC ATC AAA AGC CAT GAA GTA CAT TGG GAG      480
Ser Gln Gly Val Ile Leu Pro Ile Lys Ser His Glu Val His Trp Glu
145                 150                 155                 160

ACC TGC AGG ACA GTG CCC TTC AGC CAG ACT ATA ACC CAC GAA GGC TGT      528
Thr Cys Arg Thr Val Pro Phe Ser Gln Thr Ile Thr His Glu Gly Cys
                165                 170                 175

GAA AAA GTA GTT GTT CAG AAC AAC CTT TGC TTT GGG AAA TGC GGG TCT      576
Glu Lys Val Val Val Gln Asn Asn Leu Cys Phe Gly Lys Cys Gly Ser
            180                 185                 190

GTT CAT TTT CCT GGA GCC GCG CAG CAC TCC CAT ACC TCC TGC TCT CAC      624
Val His Phe Pro Gly Ala Ala Gln His Ser His Thr Ser Cys Ser His
        195                 200                 205

TGT TTG CCT GCC AAG TTC ACC ACG ATG CAC TTG CCA CTG AAC TGC ACT      672
Cys Leu Pro Ala Lys Phe Thr Thr Met His Leu Pro Leu Asn Cys Thr
    210                 215                 220

GAA CTT TCC TCC GTG ATC AAG GTG GTG ATG CTG GTG GAG GAG TGC CAG      720
Glu Leu Ser Ser Val Ile Lys Val Val Met Leu Val Glu Glu Cys Gln
225                 230                 235                 240

TGC AAG GTG AAG ACG GAG CAT GAA GAT GGA CAC ATC CTA CAT GCT GGC      768
Cys Lys Val Lys Thr Glu His Glu Asp Gly His Ile Leu His Ala Gly
                245                 250                 255

TCC CAG GAT TCC TTT ATC CCA GGA GTT TCA GCT TGA                      804
Ser Gln Asp Ser Phe Ile Pro Gly Val Ser Ala  *
            260                 265
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 268 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met His Leu Leu Leu Phe Gln Leu Leu Val Leu Leu Pro Leu Gly Lys
 1               5                  10                  15

Thr Thr Arg His Gln Asp Gly Arg Gln Asn Gln Ser Ser Leu Ser Pro
            20                  25                  30

Val Leu Leu Pro Arg Asn Gln Arg Glu Leu Pro Thr Gly Asn His Glu
        35                  40                  45

Glu Ala Glu Glu Lys Pro Asp Leu Phe Val Ala Val Pro His Leu Val
    50                  55                  60
```

```
Ala Thr Ser Pro Ala Gly Glu Gly Gln Arg Gln Arg Glu Lys Met Leu
 65                  70                  75                  80

Ser Arg Phe Gly Arg Phe Trp Lys Lys Pro Glu Arg Glu Met His Pro
                 85                  90                  95

Ser Arg Asp Ser Asp Ser Glu Pro Phe Pro Pro Gly Thr Gln Ser Leu
            100                 105                 110

Ile Gln Pro Ile Asp Gly Met Lys Met Glu Lys Ser Pro Leu Arg Glu
        115                 120                 125

Glu Ala Lys Lys Phe Trp His His Phe Met Phe Arg Lys Thr Pro Ala
    130                 135                 140

Ser Gln Gly Val Ile Leu Pro Ile Lys Ser His Glu Val His Trp Glu
145                 150                 155                 160

Thr Cys Arg Thr Val Pro Phe Ser Gln Thr Ile Thr His Glu Gly Cys
                165                 170                 175

Glu Lys Val Val Val Gln Asn Asn Leu Cys Phe Gly Lys Cys Gly Ser
            180                 185                 190

Val His Phe Pro Gly Ala Ala Gln His Ser His Thr Ser Cys Ser His
        195                 200                 205

Cys Leu Pro Ala Lys Phe Thr Thr Met His Leu Pro Leu Asn Cys Thr
    210                 215                 220

Glu Leu Ser Ser Val Ile Lys Val Val Met Leu Val Glu Glu Cys Gln
225                 230                 235                 240

Cys Lys Val Lys Thr Glu His Glu Asp Gly His Ile Leu His Ala Gly
                245                 250                 255

Ser Gln Asp Ser Phe Ile Pro Gly Val Ser Ala  *
            260                 265
```

# INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

**A.** The indications made below relate to the microorganism referred to in the description on page 5, line s 28-31.

**B. IDENTIFICATION OF DEPOSIT** Further deposits are identified on an additional sheet ☐

Name of depositary institution

American Type Culture Collection

Address of depositary institution *(including postal code and country)*

12301 Parklawn Drive, Rockville, Maryland, 20852, United States of America

| Date of deposit | Accession Number |
|---|---|
| 11 March 1997 | ATCC 98347 |

**C. ADDITIONAL INDICATIONS** *(leave blank if not applicable)* This information is continued on an additional sheet ☐

**D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE** *(if the indications are not for all designated States)*

**E. SEPARATE FURNISHING OF INDICATIONS** *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only

☑ This sheet was received with the international application

Authorized officer

Signed: Hosted
International Division RO/US

For International Bureau use only

☐ This sheet was received by the International Bureau on:

Authorized officer

Form PCT/RO/134 (July 1992)

What is claimed is:

1. An isolated DNA sequence comprising a DNA sequence selected from the group consisting of:

(a) nucleotides from nucleotides beginning at # 1, 52, 55, 58, 61, 64, 67, 70, 73, 121, 256, 259, 262, 265, 268, 271 or 484 and ending at #723 or 801 of SEQ ID NO:7; and

(b) sequences which hybridize to (a) under stringent hybridization conditions and encode a protein which exhibits *cerberus* activity.

2. An isolated DNA sequence comprising a DNA sequence selected from the group consisting of:

(a) nucleotides encoding amino acids starting with amino acids beginning at # 1, 18 to 25, 41, 85 to 91 or 162 and ending at #241 or 267 of SEQ ID NO: 8; and

(b) sequences which hybridize to (a) under stringent hybridization conditions and encode a protein which exhibits *cerberus* activity.

3. A vector comprising a DNA molecule of claim 1 in operative association with an expression control sequence therefor.

4. A vector comprising a DNA molecule of claim 2 in operative association with an expression control sequence therefor.

5. A host cell transformed with the vector of claim 3.

6. A host cell transformed with the vector of claim 4.

7. An isolated DNA molecule comprising a DNA sequence selected from the group consisting of:

(a) nucleotide #268 to #801 of SEQ ID NO:2; and

(b) naturally occurring allelic sequences and equivalent degenerative codon sequences of (a).

8. A vector comprising a DNA molecule of claim 7 in operative association with an expression control sequence therefor.

9. A host cell transformed with the vector of claim 8.

10. An isolated DNA molecule encoding mammalian *cerberus* protein, said DNA molecule comprising nucleotide #268 to #801 of SEQ ID NO:7.

11. An isolated DNA molecule according to claim 10, further comprising a nucleotide sequence encoding a suitable propeptide 5' to and linked in frame to the DNA coding sequence.

12. A vector comprising a DNA molecule of claim 11 in operative association with an expression control sequence therefor.

13. A host cell transformed with the vector of claim 12.

14. A method for producing purified mammalian *cerberus* protein, said method comprising the steps of:

(a) culturing a host cell transformed with a DNA sequence according to claim 1, comprising a nucleotide sequence encoding mammalian *cerberus* protein; and

(b) recovering and purifying said mammalian *cerberus* protein from the culture medium.

15. A method for producing purified mammalian *cerberus* protein said method comprising the steps of:

(a) culturing a host cell transformed with a DNA sequence according to claim 2, comprising a nucleotide sequence encoding mammalian *cerberus* protein; and

(b) recovering and purifying said mammalian *cerberus* protein from the culture medium.

16. A method for producing purified mammalian *cerberus* protein said method comprising the steps of:

(a) culturing a host cell transformed with a DNA sequence according to claim 10, comprising a nucleotide sequence encoding mammalian *cerberus* protein; and

(b) recovering and purifying said mammalian *cerberus* protein from the culture medium.

17. A purified mammalian *cerberus* polypeptide comprising an amino acid sequence according to SEQ ID NO:8.

18. A purified mammalian *cerberus* protein produced by the steps of

(a) culturing a cell transformed with a DNA comprising the nucleotide sequence from nucleotide #268 to #801 as shown in SEQ ID NO:7; and

(b) recovering and purifying from said culture medium a protein comprising the amino acid sequence from amino acid #90 to amino acid #267 as shown in SEQ ID NO:8.

19. A composition comprising a therapeutic amount of at least one mammalian *cerberus* polypeptide according to claim 18.

20. A method for altering the regulation of neuronal genes in a patient in need of same comprising administering to said patient an effective amount of the composition of claim 19.

21. A purified mammalian *cerberus* protein comprising the amino acid sequence from amino acid #1 to #267 of SEQ ID NO:8.

22. Antibodies to a purified mammalian *cerberus* protein according to claim 21.

23. A purified mammalian *cerberus* protein comprising the amino acid sequence from amino acid #90 to #267 of SEQ ID NO:2.

24. Antibodies to a purified mammalian *cerberus* protein according to claim 23.

25. A purified mammalian *cerberus* protein produced by the steps of

(a) culturing a cell transformed with a DNA comprising the nucleotide sequence from nucleotide #52 to #801 as shown in SEQ ID NO:7; and

(b) recovering and purifying from said culture medium a protein comprising the amino acid sequence from amino acid #18 to amino acid #267 as shown in SEQ ID NO:8.

# INTERNATIONAL SEARCH REPORT

Int. tional Application No
PCT/US 98/11462

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/475 C07K16/22 A61K38/18

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K A61K

Documentation searched other than minimumdocumentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | DATABASE EMBL - EMEST13<br>Entry MMAA20122, Acc.No. AA120122, 21 November 1996<br>MARRA, M. ET AL.: "mn32d09.r1 Beddington mouse embryonic region Mus musculus cDNA clone 538769 5'."<br>XP002072013<br>cited in the application<br>see the whole document<br>---<br>-/-- | 1-18,21, 23,25 |

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publicationdate of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of theinternational search | Date of mailing of the international search report |
|---|---|
| 2 September 1998 | 09/09/1998 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Smalt, R |

Form PCT/ISA/210 (second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

In ational Application No
PCT/US 98/11462

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | BOUWMEESTER T ET AL: "CERBERUS IS A HEAD-INDUCING SECRETED FACTOR EXPRESSED IN THE ANTERIOR ENDODERM OF SPEMANN'S ORGANIZER" NATURE, vol. 382, 15 August 1996, pages 595-601, XP002066227 cited in the application see the whole document | 1-18,21, 23,25 |
| P,X | BIBEN, C. ET AL.: "Murine cerberus homologue mCer-1: A candidate anterior patterning molecule." DEVELOPMENTAL BIOLOGY, vol. 194, 15 February 1998, pages 135-151, XP002072011 See the whole document, particularly p. 139, left-hand column. | 1-16,23 |
| P,X | BELO, J.A. ET AL.: "Cerberus-like is a secreted factor with neutralizing activity expressed in the anterior primitive endoderm of the mouse gastrula." MECHANISMS OF DEVELOPMENT, vol. 68, November 1997, pages 45-57, XP002072012 see the whole document | 1-16,23 |
| A | LEMAIRE P ET AL: "The vertebrate organizer: structure and molecules" TRENDS IN GENETICS, vol. 12, no. 12, December 1996, page 525-531 XP004071057 see the whole document | |

1

# INTERNATIONAL SEARCH REPORT

...ternational application No.
PCT/US 98/11462

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although claim(s) 20
is(are) directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.